CLAIMS:

1. A speech recognition apparatus comprising:

means for receiving the input signal;

means for processing the received signal to generate an energy signal which varies with local energy within the received signal;

means for filtering said energy signal to remove energy variations which have a frequency below a predetermined frequency;

means for detecting the presence of speech in said input signal using said filtered energy signal; and

means for comparing the detected speech with stored reference models to provide a recognition result.

2. An apparatus according to claim 1, wherein said filtering means is operable to remove energy variations which have a frequency above a predetermined frequency.

3. An apparatus according to claim 2, wherein said filter means is operable to filter out energy variations below 2Hz and above 10Hz.

4. An apparatus according to claim 2, wherein said filter means is operable to pass energy variations which

have a frequency of approximately 4Hz.

5. An apparatus according to claim 1, wherein said detecting means is operable to compare said filtered energy signal with a predetermined threshold and to detect the presence of speech in dependence upon the result of said comparison.

6. An apparatus according to claim 1, wherein said processing means is operable to divide the input speech signal into a number of successive time frames and to determine the energy of the input signal in each of said time frames to generate said energy signal.

7. An apparatus according to claim 6, comprising modulation power determination means for determining the modulation power of the filtered signal within a predetermined frequency band.

8. An apparatus according to claim 7, wherein said filtering means and said modulation power determining means are operable to filter and determine the modulation power in discrete portions of said energy variation signal.

9. An apparatus according to claim 8, wherein said filtering means and said power modulation determining means are formed by a discrete Fourier transform means which is operable to determine the first non-DC coefficient of a discrete Fourier transform of each discrete portion of said energy variation signal.

10. A speech recognition apparatus comprising:

means for receiving a sequence of input frames each representative of a portion of an input signal;

means for processing each frame in the received sequence of frames to generate a sequence of energy values indicative of the local energy within the representative signal;

means for filtering said sequence of energy values to remove energy variations which have a frequency below a predetermined frequency;

means for detecting the presence of speech in said input signal using said filtered energy values; and

means for comparing the detected speech with stored reference models to provide a recognition result.

11. An apparatus according to claim 10, further comprising means for determining the boundary between a speech containing portion and a background noise

containing portion in said input signal.

12. An apparatus according to claim 11, wherein said boundary determining means is operable for determining the likelihood that said boundary is located at each of a plurality of possible locations within said energy signal and means for determining the location which has the largest likelihood associated therewith.

13. An apparatus for determining the location of a boundary between a speech containing portion and a background noise containing portion in an input speech signal, the apparatus comprising:

means for receiving the input signal;

means for processing the received signal to generate an energy signal indicative of the local energy within the received signal;

means for determining the likelihood that said boundary is located at each of a plurality of possible locations within said energy signal; and

means for determining the location of said boundary using said likelihoods determined for each of said possible locations.

14. An apparatus according to claim 13, wherein said

likelihood determining means is operable to determine the likelihood that said boundary is located at each of said possible locations by: (i) comparing a portion of the energy signal on one side of the current location with a model representative of the energy in background noise; (ii) comparing the portion of the energy signal on the other side of the current location with a model representative of the energy within speech; and (iii) combining the results of said comparisons to determine a likelihood for the current possible location.

15. An apparatus according to claim 13, comprising speech detection means which is operable to process said received signal and to identify when speech is present in the received signal, and wherein said likelihood determining means is operable to determine said likelihoods in the received signal when said speech detecting means detects speech within the received signal.

16. An apparatus according to claim 13, further comprising means for filtering said energy signal to remove energy variations which have a frequency below a predetermined frequency.

17. An apparatus according to claim 16, wherein said filter means is operable to filter out energy variations below 1Hz.

18. An apparatus according to claim 13, wherein said processing means is operable to divide the input speech signal into a number of successive time frames and to determine the energy of the input signal in each of said time frames to generate a discrete energy signal.

19. An apparatus according to claim 16, wherein said filter means is operable to output a number of discrete samples representing said filtered energy signal.

20. An apparatus according to claim 19, wherein said likelihood determining means is operable to determine said likelihood for each of said discrete filtered energy values.

21. An apparatus according to claim 13, wherein said boundary is at the beginning or at the end of a speech containing portion of said received signal.

22. An apparatus according to claim 14, wherein said models are statistical models.

23. An apparatus according to claim 22, wherein said models are based on Laplacian statistics.

24. An apparatus according to claim 22, wherein said speech model is an auto-regressive model.

25. A speech recognition method comprising the steps of:

receiving the input signal;

processing the received signal to generate an energy signal which varies with local energy within the received signal;

filtering said energy signal to remove energy variations which have a frequency below a predetermined frequency;

detecting the presence of speech in said input signal using said filtered energy signal; and

comparing the detected speech with stored reference models to provide a recognition result.

26. A method according to claim 25, wherein said filtering step removes energy variations which have a frequency above a predetermined frequency.

27. A method according to claim 26, wherein said filter step filters out energy variations below 2Hz and above

10Hz.

28. A method according to claim 26, wherein said filter step passes energy variations which have a frequency of approximately 4Hz.

29. A method according to claim 25, wherein said detecting step compares said filtered energy signal with a predetermined threshold and detects the presence of speech in dependence upon the result of said comparison.

30. A method according to claim 25, wherein said processing step divides the input speech signal into a number of successive time frames and determines the energy of the input signal in each of said time frames to generate said energy signal.

31. A method according to claim 30, comprising the step of determining the modulation power of the filtered signal within a predetermined frequency band.

32. A method according to claim 31, wherein said filtering step and said modulation power determining step are operable to filter and determine the modulation power in discrete portions of said energy variation signal.

33. A method according to claim 32, wherein said filtering step and said power modulation determining step determine the first non-DC coefficient of a discrete Fourier transform of each discrete portion of said energy variation signal.

34. A speech recognition method comprising the steps of:

receiving a sequence of input frames each representative of a portion of an input signal;

processing each frame in the received sequence of frames to generate a sequence of energy values indicative of the local energy within the representative signal;

filtering said sequence of energy values to remove energy variations which have a frequency below a predetermined frequency;

detecting the presence of speech in said input signal using said filtered energy values; and

comparing the detected speech with stored reference models to provide a recognition result.

35. A method according to claim 34, further comprising the step of determining the boundary between a speech containing portion and a background noise containing portion in said input signal.

36. A method according to claim 35, wherein said boundary determining step determines the likelihood that said boundary is located at each of a plurality of possible locations within said energy signal and determines the location which has the largest likelihood associated therewith.

37. A method of determining the location of a boundary between a speech containing portion and a background noise containing portion in an input speech signal, the method comprising the steps of:

receiving the input signal;

processing the received signal to generate an energy signal indicative of the local energy within the received signal;

determining the likelihood that said boundary is located at each of a plurality of possible locations within said energy signal; and

determining the location of said boundary using said likelihoods determined for each of said possible locations.

38. A method according to claim 37, wherein said likelihood determining step determines the likelihood that said boundary is located at each of said possible

locations by: (i) comparing a portion of the energy signal on one side of the current location with a model representative of the energy in background noise; (ii) comparing the portion of the energy signal on the other side of the current location with a model representative of the energy within speech; and (iii) combining the results of said comparisons to determine a likelihood for the current possible location.

39. A method according to claim 37, comprising a speech detection step which processes said received signal and identifies when speech is present in the received signal, and wherein said likelihood determining step determines said likelihoods in the received signal when said speech detecting step detects speech within the received signal.

40. A method according to claim 37, further comprising the step of filtering said energy signal to remove energy variations which have a frequency below a predetermined frequency.

41. A method according to claim 40, wherein said filtering step filters out energy variations below 1Hz.

42. A method according to claim 37, wherein said

processing step divides the input speech signal into a number of successive time frames and determines the energy of the input signal in each of said time frames to generate a discrete energy signal.

43. A method according to claim 40, wherein said filtering step outputs a number of discrete samples representing said filtered energy signal.

44. A method according to claim 43, wherein said likelihood determining step determines said likelihood for each of said discrete filtered energy values.

45. A method according to claim 37, wherein said boundary is at the beginning or at the end of a speech containing portion of said received signal.

46. A method according to claim 38, wherein said models are statistical models.

47. A method according to claim 46, wherein said models are based on Laplacian statistics.

48. A method according to claim 46, wherein said speech model is an auto-regressive model.

49. A computer readable medium storing computer executable process steps for controlling a processor to carry out a speech recognition method, the process steps comprising the steps of:

receiving the input signal;

processing the received signal to generate an energy signal which varies with local energy within the received signal;

filtering said energy signal to remove energy variations which have a frequency below a predetermined frequency;

detecting the presence of speech in said input signal using said filtered energy signal; and

comparing the detected speech with stored reference models to provide a recognition result.

50. A computer readable medium storing computer executable process steps for controlling a processor to implement a method of detecting speech with an input signal, the process steps comprising the steps of:

receiving the input signal;

processing the received signal to generate an energy signal indicative of the local energy within the received signal;

determining the likelihood that said boundary is

located at each of a plurality of possible locations within said energy signal; and

determining the location of said boundary using said likelihoods determined for each of said possible locations.

51. Computer executable process steps for controlling a processor to implement a speech recognition method, the process steps comprising the steps of:

receiving the input signal;

processing the received signal to generate an energy signal which varies with local energy within the received signal;

filtering said energy signal to remove energy variations which have a frequency below a predetermined frequency;

detecting the presence of speech in said input signal using said filtered energy signal; and

comparing the detected speech with stored reference models to provide a recognition result.

52. Computer executable process steps for controlling a processor to implement a method of detecting the presence of speech with an input signal, the process steps comprising the steps of:

receiving the input signal;

processing the received signal to generate an energy signal indicative of the local energy within the received signal;

determining the likelihood that said boundary is located at each of a plurality of possible locations within said energy signal; and

determining the location of said boundary using said likelihoods determined for each of said possible locations.